

DOCUMENT RESUME

ED 133 338

TM 005 715

AUTHOR Scheuerman, Karl
TITLE An Inventory of Adapting Variables.
PUB DATE [Apr 76]
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS Curriculum Evaluation; Diagnostic Tests; Individual Differences; *Individualized Instruction; *Instructional Materials; Mastery Tests; *Measurement Techniques; Predictive Validity; *Student Needs

ABSTRACT

Three quantitative measures which assess instructional material in terms of the rationale of adapting are reported. The first index is termed the consequence ratio, which evaluates diagnostic items in terms of how effectively time is used. The second measure, termed predictive validity, evaluates the accuracy of diagnostic test predictions. The third index, discriminability, is a measure of how well the diagnostic test reflects individual differences. These measures have been included in a large field study which will assess the effectiveness of adapting techniques in meeting the needs of individual students. It is hoped, that by studying the actual parameters which reflect the rationale for adapting, a clearer picture of the future of adapting techniques in education will develop. For material to be truly adaptive it must score well on each of the measures. In the forthcoming field study these measures will be applied to a broad range of adaptive curricula. An analysis will reveal the contribution that adaptive features make to year end achievement test scores. (RC)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

ED133338

AN INVENTORY OF ADAPTING VARIABLES

Karl Scheuerman

Learning Research and Development Center

University of Pittsburgh

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

TM005 715

An important question to both the designers and evaluators of educational materials is whether individualization is a worthwhile technique (Holland, 1975.) Few would disagree that students have differing needs in attaining their educational goals. Yet the value of individualization techniques has never been clearly demonstrated. Part of the problem in evaluation is that until recently there have been no quantitative measures of individualization. Instructional materials which on the surface appear to be highly adaptive may not upon analysis, actually adapt to the student's needs.

Proving the value of adaptive materials must occur in two stages. First it is necessary to develop quantitative indices which reflect the rationale of adapting. Only when these measures have been developed is it possible to evaluate the usefulness of adapting techniques in terms of educational outcomes.

This paper will report three quantitative measures which assess instructional material in terms of the rationale of adapting. These measures have been included in a large scale field study which will assess the effectiveness of adapting techniques in meeting the needs of individual students (Cooley & Leinhardt, 1975.) It is hoped, that by studying the actual parameters which reflect the rationale for adapting, a clearer picture of the future of adapting techniques in education will develop.

In preparing good adaptive materials over the past decade the principles of adapting have remained implicit. The development of quantitative measures makes these principles explicit.

In the present context, individualization means that individual differences in student needs are diagnosed by adaptive features of the curriculum in an attempt to prescribe materials necessary to meet the objectives. To efficiently attain the required skills, students must come in contact with only those materials which will develop the appropriate skills.

The following are three indices of the quality of the adapting features of a curriculum in achieving this goal of efficient education. The first index is termed the consequence ratio, which evaluates diagnostic items in terms of how effectively time is used. One obvious purpose of diagnostic items is to save students' time. However, testing itself takes time. It would be a waste of time to take a lengthy test which, if passed, allowed the student to skip only a short section of teaching material. The ideal situation is one in which a short test, if passed, saved a great deal of time by allowing the student to bypass a lengthy section of teaching material.

Mathematically, the consequence ratio is the ratio of the consequence time to total time. The consequence time is the amount of time which the student would spend if the diagnostic test were failed. The total time is the sum of testing time and consequence time. In most cases, the consequence of failing a diagnostic test is that the student must take some further teaching material. However, this is not always the case. For instance, the consequence of failing a placement test may include requiring the student to take other pretests and post-tests in addition to the teaching material. Another way

of looking at the consequence time is as the total potential time which a diagnostic test, if passed, can save the student.

The consequence ratio varies between 0 and 1.00. In the ideal situation, when the test time is very short and the consequence time great, the index approaches 1.0. However when the consequence time equals the test time the ratio falls to .50. Such a consequence ratio indicates that, one student taking the test can, at best, break even. Ratios less than .50 indicate that diagnostic tests should be dropped because they waste more time than they save.

The second measure, called predictive validity, evaluates the accuracy of diagnostic test predictions. In adaptive curricula, the student who fails a diagnostic test is prescribed subsequent teaching material frequently followed by a mastery test. Both the diagnostic test and the mastery test purport to measure the same behaviors. Therefore the diagnostic test should be a good predictor of performance on the mastery test if the student is not given any intervening teaching material. This suggests a way of measuring the predictive validity of the diagnostic test.

Slide 1 illustrates four possible outcomes when the diagnostic and mastery tests are given without intervening instruction.

Insert Slide 1 here

When a student passes the diagnostic test and, without intervening instruction, also passes the mastery test, the diagnostic

test has predicted correctly or "hit." Similarly, if a student fails both tests, the diagnostic test has again predicted correctly, or hit. The diagnostic test predicts incorrectly or "misses" when the student either passes the diagnostic test and fails the mastery test, or fails the diagnostic test and passes the mastery test.

The predictive validity ratio is the ratio of correct decisions to total decisions, or the ratio of hits to the sum of hits and misses. The predictive validity ratio varies between 0 and 1.0. A ratio of 1.0 indicates that the diagnostic test has perfectly predicted performance on the mastery test and is therefore a highly valid test. When the ratio falls to .50 the diagnostic test predicts student performance as well as if a coin had been flipped. Such a test is worthless as a diagnostic tool.

The third index, discriminability, is a measure of how well the diagnostic test reflects individual differences. A diagnostic test which all students pass, or all students fail reflects no individual differences and does not prescribe differential treatments. The discriminability ratio of a diagnostic test is the ratio of the number of students who passed or failed the test (whichever is smaller) to the total number of students taking the test.

Calculated in this way, the discriminability ratio varies between 0.0 and .50. If the value is zero, meaning all students passed or all failed the diagnostic test, the test has revealed no individual differences and has not accommodated to individual needs. A discriminability ratio of zero indicates a useless diagnostic test since the prescription

for all students is identical. These indices are measures of three interrelated variables involved in the adequacy of adaptiveness of individualized teaching materials. To illustrate use of the measures, and the interrelationship of the underlying variables, I will next discuss how the measures were applied to two different adaptive programs.

The Flexer and Flexer Programmed Reviews of Mathematics series represents a typical use of diagnostic testing. In this series of six short review booklets, each teaching section has a diagnostic placement test. In the pre-publication version of the program, a single item served as the placement test for each section. Passing this item allowed the student to skip the section. However, in the published version each section was preceded by a placement test varying from five to ten items in length. Students were allowed to skip the subsequent section if they answered no more than one of these items incorrectly. We attempted to determine the values of the three adapting variables for both versions of the program.

The published version was evaluated first. Twenty-eight introductory psychology students, typical of the target population for that program, took the eight-item pretest for the first section of the logarithms unit. Each student then retook the placement test one week later without any intervening teaching material. Similarly 10 students took all three placement tests for the fraction unit and retook the test one week later. This test and retest procedure was necessary to measure predictive validity since no mastery test was included in the program. Each diagnostic test decision and the test time was

recorded for each student. Slide 2 is a summary of the results for the logarithms section.

Insert Slide 2 here

The validity ratio of .93 is quite high. The consequence ratio was calculated using teaching times published by Flexer and Flexer. Keeping in mind that a consequence ratio of .50 is the break even point, the calculated ratio of .88 is only fair. Of great interest is the surprisingly low discriminability ratio of only .07, indicating that 93% of the students tested needed to review this section of the logarithms unit.

Slide 3 reveals that the values of the measures for the fractions unit were similar to those for logarithms. The validity ratio was reasonable at .83 and the consequence ratio of .82 was also fair.

Insert Slide 3 here

The discriminability ratio of .17 was low. The poor discriminability ratios for both units exemplify a point. While both of these units have surface characteristics of individualization, neither unit adequately detects and adapts to individual differences. The placement tests tend to uniformly prescribe the remedial material. The program is not very adaptive. Given these rather disappointing results, we

decided to try to evaluate the original, single item diagnostic test.

The fraction tests were re-administered to another set of ten psychology students. It was possible to identify a single test item in each of the three fractions placement tests which was like that used in the original pre-publication version. Using these items, both the short and long forms of the tests were evaluated and compared.

Slide 4 shows the results for the long form of the fractions placement test. With a validity of .83, a consequence ratio of .83

Insert Slide 4 here

and a discriminability ratio of .17, these results are almost identical with the first evaluation. Slide 5 demonstrates the effects of giving the one item form of the placement test. The validity dropped to .73,

Insert Slide 5 here

which is approximately half way between perfect and chance prediction. On the other hand the discriminability ratio rose to the quite acceptable level of .40. And as a result of the shorter test the consequence ratio increased to an excellent .96.

There are at least two possible reasons for change in the discriminability ratio. First, with less valid tests, higher discriminability ratios are expected. If, for example, decision to skip remedial

material were made by the flip of a coin, the discriminability ratio would be a perfect .50, since exactly half the students would pass the test. The second reason is a possible change in the criterion. If the pretest criterion is lowered, more students will pass. Thus, if, for example, students were allowed two errors instead of one on the long form of the placement test, a higher discriminability ratio would be expected.

The second example selected to illustrate the application of the adapting variables is the analysis of Chapter 5 of The Arithmetic of Computers published by Crowder. This analysis is described because the program represents a different style of individualization from the straightforward placement test technique used by Flexer and Flexer.

The structure of this program is typical of Crowder's "intrinsic" programming technique. Each of the seven mainline items occurs on a single page of the text which has some material to be read, followed by a single item multiple choice test with two to four alternatives. Each alternative is followed by instructions to turn to a particular page in the text. If the multiple choice question is answered correctly the student is instructed to turn or "branch" to the next mainline item. Each incorrect answer is followed by instructions to turn to a page containing remedial material. When the remedial page has been completed the student is then returned to the mainline item page to retake the diagnostic test. It is possible to make more than one remedial excursion if the test is failed again. In general, each incorrect alternative of the multiple choice test diagnoses a particular type of remedial material. The evaluation was conducted with 9

college students. Each student followed the program as it was intended to be used. However, after a student completed a diagnostic test, progress was interrupted from 6 to 24 hours, the test was then retaken and the student resumed work, branching to any remedial material if necessary. This process was continued until all seven mainline test items had been taken.

To obtain a good, in-use record of testing and teaching times, a separate set of subjects took the program in the normal manner without interruption while an observer recorded testing and teaching times.

A summary of the results is presented in Slide 6. Most of the

Insert Slide 6 here

tests were passed yielding a high validity ratio of .97 and a low discriminability ratio of .12. This validity value may be artificially high since it was calculated by a rest-retest method using only a single-item with a multiple choice format.

The values for validity and discriminability are quite similar to those encountered in the Flexer and Flexer published program. However the consequence ratio is only .75 in contrast with the higher values observed in the previous example. The "intrinsic" programming technique may be the culprit. Having completed the remedial material, the student is again cycled through the mainline item including both teaching and testing material. Much of the consequence time incurred by failing

a diagnostic test is accounted for by page flipping and retesting.

Here is a good example of how the consequence ratio can be improved simply by making the consequence of a diagnostic item longer. However, very little of the consequence is actually teaching or remedial material. This inefficient page flipping and retesting tends to inflate the consequence ratio. There was not a single case of a student failing the diagnostic test on the second pass. If this retest time were removed, the consequence ratio would fall to .65, only .15 above the break even point. It would fall even lower if page flipping were removed.

It is difficult to guess what strategy if any might reclaim this program. Individualized programs which exercise fine-grain control over the student behavior through the use of diagnostic tests inevitably run into difficulties of the sort just described. Excessive use of diagnostics and elaborate routing strategies are usually unjustified. A diagnostic item which works efficiently has a large consequence which is mainly teaching material.

In summary, use of these quantitative indices allows us to distinguish between programs with features that superficially seem adaptive and those that really implement the theory of adapting. Neither of the program segments evaluated meet all the criteria necessary for successfully adapting to individual needs. For material to be truly adaptive it must score well on each of the measures. In the field study mentioned previously these measures will be applied to a broad range of adaptive curricula. An analysis will reveal the contribution that adaptive features make to year end achievement test scores.

References

Coeley, W. W., & Leinhardt, G. Design for the individualized instruction study: A study of the effectiveness of individualized instruction in the teaching of reading and mathematics in compensatory education programs. Final report. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, October 1975.

Crowder, N. A. The arithmetic of computers. (Vol. 1). Santa Barbara, Cal.: USI Western Design, 1960.

Flexer, R. J., & Flexer, A. S. Programmed reviews of mathematics. New York: Harper and Row, 1967.

Holland, J. G. Variables in adaptive decisions in individualized instruction. Educational Psychologist, in press.

APPENDIX

Sample Decision Table

		Criterion Test	
		Pass	Fail
Diagnostic Test	Pass	hit	miss
	Fail	miss	hit

$$\text{Predictive Validity Ratio} = \frac{\text{hits}}{\text{hits} + \text{misses}} = \frac{\text{hits}}{\text{total decisions}}$$

Slide 1

Flexer & Flexer, Logarithms

Validity and Discriminability

(1 Decision x 28 Subjects = 28 Total Decisions)

		Criterion	
		P	F
Diagnostic	P	2	0
	F	2	24

$$\text{Validity} = \frac{\text{Hits}}{\text{Hits \& Misses}} = \frac{26}{28} = .93$$

$$\text{Discriminability} = \frac{\text{Passes}}{\text{Passes \& Failures}} = \frac{2}{28} = .07$$

Consequence

$$\frac{\text{Teaching Time in Minutes}}{\text{Teaching Time \& Testing Time}} = \frac{67}{76} = .88$$

Slide 2

Flexer & Flexer, Fractions

First Evaluation - Long-Form Tests

Validity and Discriminability

(3 Decisions x 10 Subjects = 30 Total Decisions)

		Criterion	
		P	F
Diagnostic	P	5	0
	F	5	20

$$\text{Validity} = \frac{\text{Hits}}{\text{Hits \& Misses}} = \frac{25}{30} = .83$$

$$\text{Discriminability} = \frac{\text{Passes}}{\text{Passes \& Failures}} = \frac{5}{30} = .17$$

Consequence

$$\frac{\text{Teaching Time in Minutes}}{\text{Teaching Time \& Testing Time}} = \frac{150}{184} = .82$$

A. Long-form testsValidity and Discriminability

(3 Decisions x 10 Subjects = 30 Total Decisions)

		Criterion	
		P	F
Diagnostic	P	4	1
	F	4	21

$$\text{Validity} = \frac{\text{Hits}}{\text{Hits \& Misses}} = \frac{25}{30} = .83$$

$$\text{Discriminability} = \frac{\text{Passes}}{\text{Passes \& Failures}} = \frac{5}{30} = .17$$

Consequence

$$\frac{\text{Teaching Time in Minutes}}{\text{Teaching Time \& Testing Time}} = \frac{150}{181} = .83$$

Slide 4

B. Single-item testsValidity and Discriminability

(3 Decisions x 10 Subjects = 30 Total Decisions)

		Criterion	
		P	F
Diagnostic	P	15	3
	F	5	7

$$\text{Validity} = \frac{\text{Hits}}{\text{Hits \& Misses}} = \frac{22}{30} = .73$$

$$\text{Discriminability} = \frac{\text{Failures}}{\text{Passes \& Failures}} = \frac{12}{30} = .40$$

Consequence

$$\frac{\text{Teaching Time in Minutes}}{\text{Teaching Time \& Testing Time}} = \frac{150}{156} = .96$$

Slide 5

The Arithmetic of ComputersValidity and Discriminability

(9 Subjects x 7 Decisions = 63 Total Decisions*)

		Criterion	
		P	F
Diagnostic	P	52	0
	F	2	5

$$\text{Validity} = \frac{\text{Hits}}{\text{Hits \& Misses}} = \frac{57}{59} = .97$$

$$\text{Discriminability} = \frac{\text{Failures}}{\text{Passes \& Failures}} = \frac{7}{59} = .12$$

Consequence

$$\text{Consequence ratio} = \frac{\text{Consequence}}{\text{Total}} = \frac{1163}{1559} = .75$$

$$\text{Consequence ratio} = \frac{\text{Teaching}}{\text{Total}} = \frac{729}{1125} = .65$$

*Four incompleting items lowered the actual number to 59 total decisions.

Slide 6.